Enforcing Consistent Translation of German Compound Coreferences

Mascarell, Laura ; Fishel, Mark ; Korchagina, Natalia ; Volk, Martin

Abstract: Coreferences to a German compound (e.g. Nordwand) can be made using its last constituent (e.g. Wand). Intuitively, both coreferences and the last constituent of the compound should share the same translation. However, since Statistical Machine Translation (SMT) systems translate at sentence level, they both may be translated inconsistently across the document. Several studies focus on document level consistency, but mostly in general terms. This paper presents a method to enforce consistency in this particular case. Using two in-domain phrase-based SMT systems, we analyse the effects of compound coreference translation consistency on translation quality and readability of documents. Experimental results show that our method improves correctness and consistency of those coreferences as well as document readability.

Originally published at:
Mascarell, Laura; Fishel, Mark; Korchagina, Natalia; Volk, Martin (2014). Enforcing Consistent Translation of German Compound Coreferences. In: Konvens, Hildesheim, Germany, 8 October 2014 - 10 October 2014, s.n..

Enforcing Consistent Translation of German Compound Coreferences

Laura Mascarell, Mark Fishel, Natalia Korchagina and **Martin Volk**
Institute of Computational Linguistics
University of Zurich
Switzerland
{mascarell,fishel,korchagina,volk}@cl.uzh.ch

Abstract

Coreferences to a German compound (e.g. *Nordwand*) can be made using its last constituent (e.g. *Wand*). Intuitively, both coreferences and the last constituent of the compound should share the same translation. However, since Statistical Machine Translation (SMT) systems translate at sentence level, they both may be translated inconsistently across the document. Several studies focus on document level consistency, but mostly in general terms. This paper presents a method to enforce consistency in this particular case. Using two in-domain phrase-based SMT systems, we analyse the effects of compound coreference translation consistency on translation quality and readability of documents. Experimental results show that our method improves correctness and consistency of those coreferences as well as document readability.[1]

1 Introduction

Statistical Machine Translation (SMT) systems translate sentences as isolated units, ignoring document-level information (Koehn, 2009). Since this document unawareness negatively impacts the translation quality, many approaches have been proposed to introduce discourse level features in SMT.

Specifically, the issue of consistent lexical choice is our focus of attention. The one-sense-per-discourse hypothesis (Gale et al., 1992) and later the one-translation-per-discourse applied to machine translation (Carpuat, 2009) show that consistency in discourse is desirable. Some methods were then proposed to enforce consistency by applying caching (Tiedemann, 2010; Gong et al., 2011). Later, Carpuat and Simard (2012) showed that SMT systems already translate consistently, but consistency is not a good indicator of translation quality. Translation systems trained on large text collections deal with more translation choices and, therefore, they translate more inconsistently. The same study also proved that inconsistencies signal translation errors more often than consistences do. Repetition as a consequence of strict consistency enforcement is also discussed, since it is difficult to determine whether repetition is desirable or not (Carpuat and Simard, 2012). On the one hand, human translators tend to use repetition across the document. On the other hand, it may negatively affect fluency (Guillou, 2013).

Whilst all these analyses focus on a general application of consistency, in this paper, we address consistency on coreferences to a compound at document-level. We tackle a specific case in which the compound is coreferenced using its last part, proposing a method to enforce the consistent translation of those coreferences. We focus on German on the source side, since it is a language rich in compounds. For instance, considering the German-French language pair, the compound *Ostwand* (“east face” or “east wall”) in the mountaineering domain is coreferenced as *die Wand* (“the wall”). While the best French translation candidate of *Ostwand* is *face nord*, *Wand* as an isolated word is more likely to be translated as *paroi*. Here we assume that the last part of a compound and its coreferences should share the translation, which would help to identify connectedness between sentences. In our experiments, we assess consistency and its correlation with translation quality in this particular case. Although it is not clear that repetition is always desirable, does it improve readability?

In the following, section 2 gives an overview of the work related to consistency in SMT. After describing our method to enforce consistency in section 3, we detail the carried out experiments in section 4 and discuss the results in section 5.

2 Related Work

The well-known one-sense-per-discourse hypothesis (Gale et al., 1992) was later applied to machine translation as one-translation-per-discourse (Carpuat, 2009), proving that more than one translation per discourse is often due to wrong lexical choices. Based on this constraint, some studies focused on analysing consistency in SMT. Carpuat and Simard (2012) analysed how consistent is the output of SMT systems compared to human translations. They experimented with several phrase-based SMT systems trained on different conditions such as data size, domain or language pair. The work showed that SMT systems translate nearly as consistently as human translators. However, inconsistency often points to translation errors and therefore cannot be ignored. Guillou (2013) studied a different approach analysing *when* (i.e. genre) and *where* (i.e. part-of-speech) lexical consistency is desirable.

Several approaches focused on enforcing consistent lexical choice. Tiedemann (2010) proposed a cache-based model that propagates the translation of phrases across the document. However, the caching approach is sensitive to error propagation. Gong et al. (2011) extended the approach applying a dynamic, static and topic cache, where the latest keeps the error propagation problem controlled. Xiao et al. (2011) described a three steps procedure that enforces consistent translation of ambiguous words and Ture et al. (2012) introduced cross-sentence features to the translation model, achieving improvements on the Arabic-English language pair.

3 Enforcing Consistent Translation

A compound can be coreferenced by its last constituent. For example, the compound *Nordwand* ("north face"), formed by *Nord* (*X*) and *Wand* (*Y*) can be coreferenced by *Wand* alone (*Y*)[2]. The main aim of our method is to detect such cases and to enforce that last constituent *Y* to have the same translation in both *XY* and *Y*.

To consistently translate *Y*, we cache its translation and enforce it when a coreference is detected. This greedy approach is sensitive to error propagation in general; however, our method is restricted to compounds, which provide more context for a correct translation than single roots, yielding less translation variants.

In detail our method works as follows. We translate each sentence individually, caching the translation of the last part of compounds and enforcing a translation for a coreference when required. To identify compounds, first we analyse each noun with the German morphology system Gertwol (Koskeniemmi and Haapalainen, 1994), which marks the boundaries between independent morphemes (e.g. the analysis of *Ostwand* is *Ost#wand*). Next, we obtain the translation of a compound from the word alignment given by the SMT decoder. We then check at the phrase table which one of its content words is the translation of *Y*, and we cache it. For instance, considering the German compound *Bundesamt* ("federal office"), which is aligned to the French *le office fédéral* in the target side, and its coreference *Amt* ("office"), we cache the pair *Amt* and *office*. If there are several compounds sharing the last morpheme, *Y* usually will corefer to the closer one, but not necessarily, which intoduces an ambiguity problem. For instance, the noun *Wand* ("wall") in *Ostwand* ("east face") and *Felswand* ("rockface") is translated into French as *face* and *paroi*, respectively. An analysis of local context would provide better precision, but for the sake of simplicity, we assume that *Y* corefers to the last compound translated, always caching its last occurrence.

To identify *Y* as a coreference, we apply the pattern "determiner + (adjective) + *Y* lemma",

[2] Compounds can consist of more than two roots and thus also *X* and *Y*. For instance, considering the compound *Eigernordwand* ("Eiger north face"), *Y* can be either *Wand* or the compound *Nordwand*.

where the *adjective* is optional and the *determiner* is tagged as one of the following parts-of-speech: PDS (substituting demonstrative pronoun), PDAT (attributive demonstrative pronoun), PPOSS (substituting possessive pronoun), PPOSAT (attributive possessive pronoun) or ART (restricted to definite articles). Thus, *die prächtige Fahrt* and *diesen Grat* are examples matching the pattern. We use the lemma of *Y* to also match examples where German cases (e.g. genitive and dative) change the form of *Y* (e.g. *Grates* is the genitive form of *Grat*). We then check that *Y* is cached and there is a compound *XY* in the four preceding sentences[3]. In case of PDAT (e.g. *diese*), we consider the whole document. PDAT is a strong coreference indicator, and we found examples having more than four sentences between the compound and its coreference. The cached translation of *Y* is then plugged into the decoder.

4 Experiments

We first conduct an analysis of compounds and coreferences automatically detected, which is mostly manually addressed by two different annotators. We then carry out all the experiments on the German-French language pair, testing different approaches to plug the translation into the decoder and to increase the coverage of our method.

The data comes from the Text+Berg corpus (Bubenhofer et al., 2013), a collection of documents from the Alpine domain, which was built as a result of digitising and processing the Swiss Alpine Club yearbooks from 1864 to 2009 (Volk et al., 2010). The sentence alignment was carried out with Bleualign (Sennrich and Volk, 2010). The test set in both manual analysis and translation task is a collection of 318 examples, that is, groups of sentences containing a compound noun and its coreferences, randomly sampled from Text+Berg data (see section 4.1).

4.1 Analysing the detected compounds and their coreferences

To evaluate how often a German compound is coreferenced using its last constituent, we automatically detect them in a German corpus consisting of roughly 1.1 million sentences from Text+Berg. Our method is the same described in section 3 to identify compounds and their coreference. We found 24,317 cases where this occurs, and to assess the effectiveness of our method at detecting a compound and its coreferences, we carried out a manual analysis on a random sample containing 318 compound-coreference pairs automatically detected. This analysis shows that 107 of these pairs are false positives, that is, the coreference is incorrectly detected, often due to the lexicalization of the compound or a number disagreement between compound and coreference. A lexicalized compound cannot be coreferenced by its last part, since its translation does not correspond to the translation of its constituents. For example, the German compounds *Zusammenarbeit* ("cooperation") and *Augenblick* ("moment") are lexicalized and thus, they cannot be corefereced by *die Arbeit* ("the work") or *der Blick* ("the view"), respectively. The disagreement in number is due to our method match the lemma of *Y* to also detect examples when the German cases change the word forms. In other less frequent cases, the detected coreference has nothing to do with the compound. For instance, in the example shown in Table 1, the pattern matches correctly the coreference *Gipfel* ("summit"), but the method fails at detecting *Schneegipfel* ("snowy summit") as the compound coreferenced. Indeed, *Gipfel* ("summit") corefers to the mountain *Königsspitze*.

The manual analysis also focuses on the correct detections, distinguishing the following most common patterns:

- The coreference is preceded by a definite article + adjective or by the demonstrative adjectives *dieser* ("this") and *jener* ("that") in all their grammatical forms.
- The compound is in genitive case and its coreference in nominative or dative case. For example, *das Tal* ("the valley") corefers to *Haupttals* ("main valley") in *Sohle des Haupttals* ("bottom of the main valley").

4.2 Enforcing translation

The translation of compounds is the first step to proceed with our method. However, compounds are often Out-Of-Vocabulary (OOV) (i.e. they do

[3]We carried out several experiments with different number of sentences. We decided to use a four sentence window, since more than four introduces too noise.

Er sah von ihr wirklich auf den obern Trafoierferner links hinunter und erblickte über mehrere ***Schneegipfel*** hinweg sein Ziel, die im Hintergrunde sich erhebende Königsspitze .	
Auf deren ***Gipfel*** grub er sich dann halbliegend in den zusammengewehten Schnee ein .	
"He looked from her to the upper Trafoierferner down to the left and saw several *snow peaks* across his goal, which is in the background of *Königsspitze*."	
"On its *summit* he dug himself in a half-lying in the snow along a wind-blown."	

Table 1: Example where the coreference to a compound was incorrectly detected.

not appear in the training corpus) and the system cannot translate them. These compounds are usually composed of frequent words in the training corpus, so we can obtain the translation of an unseen compound by splitting it into its known parts and translating them (Koehn and Knight, 2003).

We want to assess the performance of our method in both approaches (i.e. splitting compounds and not splitting them), so we build two phrase-based SMT systems *SMT-1* and *SMT-split*, where *SMT-split* performs compound splitting. Both systems are built using the standard settings (Koehn et al.,), 5-gram language model KenLM (Heafield, 2011) and GIZA++ (Och and Ney, 2003). The language model is trained on a total of 624,160 sentences (13 million target tokens) and the training set consists of 219,187 sentences and roughly 4.1/4.7 million words in German and French, respectively. The SMT systems are tuned with Minimum Error Rate Training (Och, 2003) on a development set, also from Text+Berg, consisting of 1,424 sentence pairs and approximately 31,000 tokens for each language. We expect to enforce a consistent translation in a higher number of cases with the *SMT-split* system. Furthermore, the splitting method allows us to have a one-to-one alignment between the compound constituents and their translation. Thus, we can identify the translation of the last part of the compound and cache it directly.

Once a compound *XY* is translated, and in order to enforce the correct translation of *Y*, we explore two approaches. The idea is to find out which is the best at selecting the translation candidate. The first one lets the decoder decide which is the best translation of *Y*. We first cache the translation of a compound *XY* as a translation of *Y*, and when a coreference *Y* is detected, we plug all the content words cached into the decoder, not assigning any probability to them, so by default they are 1. Then, the decoder chooses the best candidate based on translation and language model scores. Interestingly, this first approach fails in our experiments. Most of the time, the decoder takes the translation of the first constituent of the compound instead of the last one. For instance, *Wand* as a coreference of *Nordwand* (French translation: *nord face*) is enforced to be translated into *nord*. We think that if the first constituent of a compound appears more frequently in the language model, the score computed is then higher and it is then picked as the translation candidate.

In the second approach, for each content word of a compound translation, we check that it appears as a translation candidate of *Y* in the phrase table. We then cache only the one that has the highest direct phrase translation probability. We observe that by applying this method, some examples where the compound was aligned to one word in the target side due to a misalignment or lexicalization of the compound are improved. In the first approach, we consider that these examples enforce an incorrect translation to the coreference, so they are detected as false positives and discarded. However, in this second approach, the translation is enforced, since it appears as a translation candidate of *Y* in the phrase table, resulting in a better translation of the term according to the context. The second example in the Table 2 shows that the term *Fahrt* is translated into *ascension*, which is also the translation of the compound coreferenced (*Bergfahrten*).

The results in section 5 are obtained with the second approach. Moreover, we use the automatic generator from the Apertium[4] MT toolbox to generate the correct form of those cases where compound and coreference do not agree in number.

[4] www.apertium.org

Source	Die Originalauswertung wurde in den Zwischenmassstab 1:20000 reduziert, worauf das ***Bundesamt*** (trans: *office fédéral*) für Landestopographie in Aktion trat. Nur dieses ***Amt*** war in der Lage, [...]
English translation by the authors	"The original evaluation was reduced in the intermediate scale 1:20000, followed by the *Federal Office* of Topography went into action. Only this *office* was able to [...] "
SMT-1, SMT-split	que ce ***poste*** tait dans la situation, [...]
SMT-1 enf., SMT-split enf.	que de cet ***office*** tait en mesure [...]
Source	Unter den Neuen ***Bergfahrten*** (trans: *ascension*) in den Schweizeralpen ist im IV. Band der Alpen 1928 eine erste Begehung des ganzen Südostgrates von der Gemsenlücke [...] über die prächtige ***Fahrt*** geblieben.
English translation by the authors	"Among the new *hill climbing* in the Swiss Alps is mentioned in the fourth volume of the Alps 1928 a first ascent of the whole South East ridge of the Gemsenlück [...] remained about the magnificent *journey*."
SMT-1, SMT-split	par cette magnifique ***course***.
SMT-1 enf., SMT-split enf.	par cette magnifique ***ascension***.
Source	Einen Teil ihrer bergsteigerischen und wissenschaftlichen Erfolge finden unsere Mitglieder in diesem ***Quartalsheft*** (trans: *présent numéro trimestriel*) verzeichnet. Das vorliegende ***Heft*** möge daher [...]
English translation by the authors	"Our members find part reported of their mountaineering and scientific achievements in this *quarterly bulletin*. This *bulletin* may therefore [...]
SMT-1, SMT-split, SMT-1 enf.	le ***cahier*** möge donc [...]
SMT-split enf.	le présent ***numéro*** möge donc [...]
Source	Dass dies gemacht wird, zeigt das Routenbuch Clean-Begehungen, das im ***Klettergebiet*** (trans: *site d'escalade*) liegt. Wir diskutieren über die schönsten Routen im ***Gebiet***.
English translation by the authors	"That this is done, the route book shows Clean inspections, which is located in the *climbing area*. We discuss about the best tours in the *area*."
SMT-1, SMT-split	nous discutons sur les plus belles voies dans la ***région***.
SMT-1 enf., SMT-split enf.	nous discutons sur les plus belles voies du ***site***.

Table 2: Examples where our enforcing method improves the translation of the coreference. The first example shows that the enforcing method improves the translation of *Amt*. In the second example, the compound is aligned to only one word in the target side, but its coreference translation is correctly enforced and improved. In the third example, SMT-1 misaligned *Quartalsheft* to only *trimestrel*, thus the coreference is not enforced. Due to the compound splitting technique, there is one-to-one correspondence between *sheft* and *numéro*, then the coreference translation is successfully enforced. In the last example, both translations of *Gebiet* are correct, but *site* is consistent with the translation of the compound coreferenced.

5 Results

We present results on both correctness and consistency. The analysed systems are *SMT-1* and *SMT-split* with and without applying our enforcing method. The experiments are performed on the test set consisting of 211 compound-coreference pairs correctly detected. To get those results, two annotators conducted a manual analysis and annotation of the results. The agreement between them at the task of deciding "is/is not a coreference" and "is correct/wrong coreference translation" is 73.4% and 86.8%, respectively.

Automatic detection of coreferences to a compound: The precision of our method correctly detecting a coreference to a compound is 66.4% (i.e. 211 out of 318 coreferences). We only analyse the sentences detected by the method, so recall is not computed. However, our detector's approach is broad-coverage-oriented, that is, it tends to detect more false positives examples while practically avoiding false negatives.

Coverage of the method: We compute statistics on the examples where a translation is enforced in both correct and incorrect detection of compound and coreference. When we do not perfom splitting, 42.2% (i.e. 89 out of 211) of the positive examples and 27.1% (i.e. 29 out of 107) of the incorrectly detected are enforced. The remaining 57.8% of the positive examples (i.e. where no enforcing is applied) is due to OOV compounds and misalignments. Splitting significantly increases the coverage of enforced translations from 42.2% to 56.4% (i.e. 119 out of 211). The incorrectly identified coreferences have again a lower impact ratio (34.6%; 37 out of 107).

Consistency and correctness of *SMT-1*: The *SMT-1* system without enforcing translates correctly with 80.1% accuracy and 27.5% consistency (see Table 3). The German noun *Wand* is the most common example of inconsistent but correct translation in our test set. The most likely translation for this noun is *paroi* in the Text+Berg corpora. However, when *Wand* is part of a compound, it is usually translated into *face*.

Our method applied to *SMT-1* enforces a consistent translation in 89 of the cases improving the translation of six of them and 15 cases stay correct, but become consistent. For instance, at the last example in Table 2, the noun *Gebiet* ("area") is translated into *site* instead of *région* when a consistent translation is enforced, yet both translations are correct. Furthermore, a coreference to a compound stays incorrect but become consistent, increasing the value of incorrect and consistent by one (see Table 4). The remaining 67 stay unmodified, that is, *SMT-1* chooses the consistent translation for the coreference without enforcing. Thus, while the correctness is slightly raised from 80.1% to 82.9%, the consistency improves from 27.5 to 37.9%.

	Consistent: yes	Consistent: no
Correct	52	117
Incorrect	6	36

Table 3: Consistency and correctness results of the *SMT-1* system without enforcing consistency.

	Consistent: yes	Consistent: no
Correct	73	102
Incorrect	7	29

Table 4: Consistency and correctness results of the *SMT-1* system when our enforcing method is applied.

Consistency and correctness of *SMT-split*: When we perform splitting, three cases become worse, but most of the cases that are not enforced with the *SMT-1* system due to a misalignment or OOV compounds, are now enforced and improved. For instance, the third example in Table 2 shows that the translation of the German noun *Heft* is only well enforced with the splitting approach, since without splitting, the compound *Quartalsheft* is misaligned to only *trimestrel*. The *SMT-split* system without enforcing translates correctly with 82.0% accuracy and 35.1% consistency (see Table 5).

	Consistent: yes	Consistent: no
Correct	68	105
Incorrect	6	32

Table 5: Consistency and correctness results of the *SMT-split* system without enforcing consistency.

When we apply enforcing to the *SMT-split* system, the coverage is increased and more improvement is shown (Table 6). Indeed, it applies enforcing to 109 cases improving 10 of them. Although there are six consistent and incorrect cases in both Table 3 and Table 6, some of them are different. Specifically, *SMT-split* improves two of them and makes consistent another two, although both stay incorrect. The correctness rises from 82.0% to 86.7% and consistency from 35.1% to 52.1%.

	Consistent:	
	yes	no
Correct	103	80
Incorrect	6	22

Table 6: Consistency and correctness results of the *SMT-split* system when our method is applied.

	Correctness	Consistency
SMT-1	80.1%	27.5%
SMT-split	82.0%	35.1%
SMT-1 enf.	82.9%	37.9%
SMT-split enf.	86.7%	52.1%

Table 7: Overall percentages of consistency and correctness results of *SMT-1* and *SMT-split* systems, with and without applying our enforcing method.

Overall, the final effect is positive (see Table 7). Correctness rises from 80.1% to 86.7%, improving 17 examples, that is, one third of errors are fixed, and consistency from 27.5% to 52.1%.

6 Conclusions

We present a method to enforce consistent translation of coreferences to a compound, when the coreference matches with the last constituent of the compound coreferenced. We assess correctness and consistency with two systems *SMT-1* and *SMT-split*, where the latest performs compound splitting. We then evaluate how well our method performs when applied in both systems.

We also conduct a manual analysis on the source side. We detect that the demonstrative adjectives *dieser* ("this") and *jener* ("that") are strong indicators of coreference. Furthermore, compounds are often in genitive case and their coreferences in either nominative or dative. The incorrect detection of a coreference to a compound are often due to a lexicalization of the compound and number disagreement between compound and coreference. Note that we match lemmas to abstract away from morphological changes due to the German cases (e.g. genitive or dative).

Experimental results show that the Statistical Machine Translation (SMT) systems often translate correctly and consistently coreferences to a compound. However, when our method is applied, some cases are improved and there are only few cases where the translation become worse. When the translation is successfully improved, it usually enforces a more specific term in the context. Since the splitting method allows the *SMT-split* system to translate out-of-vocabulary compounds, *SMT-split* increases the number of the enforced examples, improving the translation in a higher number of cases. Finally, we point out the importance of consistency in this study. At the examples where the coreference is correct, but inconsistent, our method also enforces consistency, which helps the reader to identify connectedness between coreference and compound, improving the readability of the document.

7 Future Work

We want to extend the study testing our method with an out-of-domain system. We expect that compounds will be correctly translated, but not their coreferences. Then, our method would enforce a correct translation, improving the output of the machine translation system.

Another case of study is when the compound is coreferenced using its first constituent rather than its last. The following made-up example shows that *triples* in the phrase *the identified triples* is a coreference of *triple structures*. Note that *triple* has been nominalized in the coreference.

> *[...] to identify **triple structures** [...]*
> ***The identified triples** [...]*

Since the first constituent of a compound is often used to describe the rest of it, we want to analyse whether the compound could be coreferenced by the nominalization of its first part.

We detected also cases where the coreference is not the last part of the compound coreferenced, but a synonym instead. For example, *Nordwand* ("north face" or "north wall") can be also corefered by *dieser Mauer* ("this wall"). We want to assess how consistency impacts on these examples, where the source is already inconsistent.

Acknowledgments

The authors would like to thank Bonnie Webber for pointing out coreferenced noun compounds as a problem for machine translation. This research was supported by the Swiss National Science Foundation under grant CRSII2_147653/1 through the project "MODERN: Modelling discourse entities and relations for coherent machine translation".

References

Noah Bubenhofer, Martin Volk, David Klaper, Manuela Weibel, and Daniel West. 2013. Text+Berg-korpus (release 147_v03). Digitale Edition des Jahrbuch des SAC 1864-1923, Echo des Alpes 1872-1924 und Die Alpen 1925-2011.

Marine Carpuat and Michel Simard. 2012. The trouble with SMT consistency. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 442–449, Montréal, Canada.

Marine Carpuat. 2009. One translation per discourse. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pages 19–27, Boulder, Colorado.

William A Gale, Kenneth W Church, and David Yarowsky. 1992. One sense per discourse. In *Proceedings of the Workshop on Speech and Natural Language*, pages 233–237, Harriman, New York.

Zhengxian Gong, Min Zhang, and Guodong Zhou. 2011. Cache-based document-level Statistical Machine Translation. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 909–919, Edinburgh, UK.

Liane Guillou. 2013. Analysing lexical consistency in translation. In *Proceedings of the Workshop on Discourse in Machine Translation*, pages 10–18, Sofia, Bulgaria.

Kenneth Heafield. 2011. KenLM: Faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 187–197, Edinburgh, Scotland.

Philipp Koehn and Kevin Knight. 2003. Empirical methods for compound splitting. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics*, volume 1, pages 187–193, Budapest, Hungary.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Edmonton, Canada.

Philipp Koehn. 2009. *Statistical Machine Translation*. Cambridge University Press, New York, USA.

Kimmo Koskeniemmi and Mariikka Haapalainen. 1994. Gertwol–lingsoft oy. *Linguistische Verifikation: Dokumentation zur Ersten Morpholympics*, pages 121–140.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51.

Franz Josef Och. 2003. Minimum error rate training in Statistical Machine Translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, volume 1, pages 160–167, Sapporo, Japan.

Rico Sennrich and Martin Volk. 2010. MT-based sentence alignment for OCR-generated parallel texts. In *The Ninth Conference of the Association for Machine Translation in the Americas*, Denver, Colorado.

Jörg Tiedemann. 2010. Context adaptation in Statistical Machine Translation using models with exponentially decaying cache. In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, Uppsala, Sweden.

Ferhan Ture, Douglas W. Oard, and Philip Resnik. 2012. Encouraging consistent translation choices. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies*, pages 417–426, Montréal, Canada.

Martin Volk, Noah Bubenhofer, Adrian Althaus, Maya Bangerter, Lenz Furrer, and Beni Ruef. 2010. Challenges in building a multilingual alpine heritage corpus. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, Valletta, Malta, may.

Tong Xiao, Jingbo Zhu, Shujie Yao, and Hao Zhang. 2011. Document-level consistency verification in Machine Translation. In *Proceedings of the 13th Machine Translation Summit*, volume 13, pages 131–138, Xiamen, China.